Atomic Clusters with Addressable Complexity

David J. Wales[1, a)]

University Chemical Laboratories, Lensfield Road, Cambridge CB2 1EW, United Kingdom

A general formulation for constructing addressable atomic clusters is introduced, based on one or more reference structures. By modifying the well depths in a given interatomic potential in favour of nearest-neighbour interactions that are defined in the reference(s), the potential energy landscape can be biased to make a particular permutational isomer the global minimum. The magnitude of the bias changes the resulting potential energy landscape systematically, providing a framework to produce clusters that should self-organise efficiently into the target structure. These features are illustrated for small systems, where all the relevant local minima and transition states can be identified, and for the low-energy regions of the landscape for larger clusters. For a 55-particle cluster it is possible to design a target structure from a transition state of the original potential, and to retain this structure in a doubly-addressable landscape. Disconnectivity graphs based on local minima that have no direct connections to a lower minimum provide a helpful way to visualise the larger databases. These minima correspond to the termini of monotonic sequences, which always proceed downhill in terms of potential energy, and we identify them as a class of biminimum. Multiple copies of the target cluster are treated by adding a repulsive term between particles with the same address to prevent excessive aggregation. By tuning the magnitude of this term it is possible to create assemblies of the target cluster corresponding to a variety of structures, including rings, chains, rotaxanes and catenanes.

[a)]dw34@cam.ac.uk

I. INTRODUCTION

To construct an operational machine we generally need to assemble a variety of components into a well-defined spatial arrangement. The experimental realisation[1] of programmed self-assembly for a structure composed of thousands of distinct building blocks has therefore generated great interest. Here the building blocks are 'DNA bricks', which can bind by hybridisation to four neighbours. The resulting assemblies are considered 'addressable', in that the different components are located in specific local environments. Understanding and developing design principles for such structures could provide a route to translation of information encoded in nanoscale building blocks into new materials with specific structure and function. Developing models that reproduce the key experimental results using the simplest possible representations is therefore an important challenge, and initial efforts for DNA bricks have already reproduced addressable assembly for as many as 1000 distinct components.[2] Recent insight from computer simulation indicates that robust self-assembly may require precise conditions for nucleation to occur. The yield for the target structure can also be improved significantly using a specific annealing protocol.[3]

Clusters of colloidal particles, functioning as pseudoatoms, also hold the prospect of designed addressable structures.[4,5] Direct observation of structure, dynamics, and thermodynamics can be obtained or inferred from optical microscopy for building blocks composed of polystyrene microspheres, inspiring complementary theoretical analysis.[6–9] Design principles have been considered for these clusters,[10] including an approach[7] based on the random energy model[11] for protein folding.[12] The latter results suggest that optimal yields for self-assembly may be

obtained when the nearest-neighbour interactions between different components are of comparable strength.[7]

In the present contribution a general framework for constructing an addressable interatomic potential is suggested, which uses the pairwise Lennard-Jones form[13] for specificity:

$$V = 4 \sum_{i<j} \epsilon_{ij} \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^{6} \right] \tag{1}$$

where r_{ij} is the distance between atoms i and j. Here the well depth is employed to produce addressable functionality. As for structure-based potentials and Gō models,[14] we introduce a bias towards a particular local minimum, in this case a specific permutational isomer of a selected stationary point. In the present work, we simply define

$$\epsilon_{ij} = \begin{cases} \epsilon, & r_{ij}^0 \leq r_c, \\ \alpha\epsilon, & r_{ij}^0 > r_c, \end{cases} \tag{2}$$

where $0 \leq \alpha \leq 1$, r_{ij}^0 is the distance between atoms i and j in the reference structure, indicated by the superscript 0, and r_c is a cutoff. Analysing the effect of α on the underlying landscape is the main objective of the present work. ϵ and σ are chosen as the units of energy and distance, and r_c was set to 1.2 for all the calculations reported below. Hence pair interactions that do not correspond to nearest neighbours in the reference structure are disfavoured as α is reduced from unity.

Biasing the potential with reference to a particular configuration has been considered before in studies of glassy systems, including structural glass formers[15,16] and spin models.[17] In these models the potential is shifted according to the overlap with a given reference configuration, defined in terms of a sum of step functions

over all distinct pairs of atoms for the atomic models. The principal difference from the present work is that the addressable potential defined above is based on a particular permutational isomer. This choice partly reflects alternative theoretical perspectives, where models based upon density are often used in treating condensed matter and glasses, while a single-particle view is usually adopted for clusters and finite systems. Overlap functions have also proved useful in studies of glass forming systems that employ pinned (frozen) particles[18–27] and cavities defined by a frozen atomic environment.[28–37]

There are some similarities between the addressable potential defined in equation (2) and elastic network models[38] designed to treat conformational changes in biomolecules.[39–56] The common theme is identification of nearest neighbours, but the network models usually implement this structure through local harmonic springs. Double-well Gō models have also been defined for biomolecules,[57–61] where switching is achieved between parameterisations biased towards distinct minima.

In this report we consider a doubly-addressable potential by assigning

$$\epsilon_{ij} = \begin{cases} \epsilon, & \min(r_{ij}^0, r_{ij}^1) \leq r_c, \\ \alpha\epsilon, & \min(r_{ij}^0, r_{ij}^1) > r_c, \end{cases} \tag{3}$$

using the minimum pair distance for two reference structures, 0 and 1. This formulation is similar in spirit to that of Murugan *et al.*,[62] who considered 'multifarious' structures composed of a variety of components. In the latter model, favourable pair interactions were included for components that needed to interact in any of a number of target structures. The capability to select alternative targets was then demonstrated using different (super)critical nucleation seeds. The po-

tential considered in the present work, defined in equation (3), defines two targets in a similar way. The focus here is on the appearance of the underlying energy landscape, especially in terms of the self-assembly characteristics defined by funnelling properties. For the doubly-addressable potential, it might be possible to select between the two alternatives using a kinetic approach based on nucleation, especially for the example considered where the underlying packing is very different. Alternatively, with some further modification, alternative structures might be selected using external parameters, such as an applied field.[63]

Experimentally, addressable self-assembly usually involves formation of multiple copies of the target structure, for example, from a solution of DNA strands. DNA-based materials have been used extensively in previous experiments, for functionalised polymer microspheres[64–66] and gold particles,[67–70] DNA scaffold and staple systems,[71,72] single-stranded tiles and bricks.[1,73,74] Alternatively, some recent advances have focused upon self-assembly of synthetic peptides.[75,76] The present model is intended to be as general as possible, and might help to guide experiments based on a variety of building blocks.

Establishing how the energy landscape depends upon parameters of the model potential for a single incarnation of the target structure provides the foundations for treating multiple copies. For a target containing N particles we can extend the formulation of the potential using the function $D(i,N) = \mathrm{Mod}(i,N) + 1$, where $\mathrm{Mod(i, N)}$ is the particle index modulo N, ranging from 0 to $N-1$. The function

$D(i,N)$ therefore produces an address label in the range 1 to N, and we choose

$$\epsilon_{ij} = \begin{cases} \epsilon, & r^0_{D(i,N)D(j,N)} \leq r_c, \\ \alpha\epsilon, & r^0_{D(i,N)D(j,N)} > r_c, \\ 0, & D(i,N) = D(j,N), \end{cases} \tag{4}$$

with an additional repulsive term between particles that share the same address:

$$V^{\text{rep}} = 4 \sum_{i<j} \epsilon^{\text{rep}} \left(\frac{\sigma}{r_{ij}} \right)^{12} \delta \left[D(i,N), D(j,N) \right], \tag{5}$$

where $\delta[k,l]$ is the Kronecker delta, equal to one if $k = l$ and zero otherwise. This formulation is invariant to exchange of particles with the same address.

Preliminary results for systems containing multiple copies of the target cluster are presented in §V. If ϵ^{rep} is too small, the target clusters can aggregate to produce conglomerates where the individual clusters are distorted or difficult to distinguish. However, as ϵ^{rep} increases, the favoured structures for the aggregates contain well-defined copies of the target cluster, assuming that we have an equal number of particles for each of the N address labels. The global minimum changes as a function of ϵ^{rep}, passing through stacked and single rings of clusters (§V).

The formulation in equation (5) was chosen because it is a simple and convenient way to tune the structures supported by the model. The aim is to understand the minimal conditions on the potential required to realise these structures. To make contact with particular experiments it will be necessary to map the interparticle forces that are under experimental control into the model, and vice versa. The increasing capabilities to tune interactions between building blocks based upon DNA and colloids suggest that it will be possible to exploit this route in future work that treats specific applications.

The principal advantage of the present model is that it enables us to analyse the organisation of the underlying energy landscape in detail, and predict changes in the likely self-assembly characteristics in terms of the strength of the competing interactions. For example, we can tune the energy scales associated with higher energy structures, where particles are not in their optimal arrangements within each N-body cluster, and the barriers between alternative packings of the clusters within aggregates. The resulting energy landscapes can be analysed in terms of local rearrangements within each cluster, and exchanges of particle with the same address between different clusters. Initial results are reported in §V.

II. METHODS

All the methodology employed in the present work for exploring the potential energy landscape and constructing kinetic transition networks[77–79] is well established, and has been described in detail before. In brief, the characterisation of each system begins with a basin-hopping global optimisation[80–82] survey, which harvests low-lying minima along with the global minimum. For the small clusters considered here, we can find all the distinct minima in this stage, and we then attempt to connect them pairwise using the doubly-nudged[83] elastic band[84,85] (DNEB) approach, which gives most, if not all, the distinct transition states after refinement of candidate structures using hybrid eigenvector-following.[86–88] Here we define transition states geometrically, as saddle points of index one, with a single negative Hessian eigenvalue.[89] The connection attempts were repeated for different DNEB parameters, such as image and iteration density, until no new transition states were obtained. Cross-relaxation checks were conducted for the

databases constructed with different values of the well-depth parameter α, as described in §III. The databases are visualised using standard constructions for disconnectivity graphs.[90,91]

For the larger clusters, containing 13 and 55 particles, we can only expect to converge the landscape in the region of the global minimum. Here we refined stationary point databases using various tools developed within the discrete path sampling[92,93] framework, as implemented in the `PATHSAMPLE` program.[94]

III. RESULTS

The energy landscapes were characterised for individual target clusters of 5, 6, 7, 13 and 55 atoms. For $5 \leq N \leq 7$ complete enumeration of all the minima and transition states should be feasible. Starting from the stationary points for $\alpha = 1$ all the permutation-inversion isomers were constructed and then relaxed for successive values of $\alpha = 0.75$, 0.5 and 0.25 using LBFGS minimisation for the minima and hybrid eigenvector-following for the transition states,[86] as described above. This sequence of α values corresponds to changing the potential in the least abrupt manner. LBFGS is a limited memory version of the quasi-Newton Broyden,[95] Fletcher,[96] Goldfarb,[97] Shanno,[98] BFGS algorithm. Permutation-inversion isomers are structures that cannot be superimposed by an overall rotation, but differ only through the arrangement of atoms of the same element or inversion of all particle coordinates through the origin.

As a cross-check, the databases obtained for each α were subsequently relaxed for all the other α values considered, and this process was repeated until no new stationary points were obtained. Relatively small maximum step sizes were em-

ployed in all the geometry optimisations to relax structures slowly. A convergence condition of 10^{-10} reduced units for the root mean square gradient was adopted, for which energy differences of 10^{-13} or more could be distinguished.

In each case results were obtained for a particular permutation-inversion isomer of the global minimum as the reference. These systems will be denoted by a $*$ superscript, e.g. LJ_{13}^{*} for an icosahedral reference geometry with 13 atoms. For LJ_{55} two additional potentials were considered, one for a transition state of O_h symmetry as the reference, denoted $\mathrm{LJ}_{55}^{O_h}$, and a doubly-addressable system with the two reference geometries corresponding to isomers of the I_h global minimum and the O_h transition state. The cuboctahedral transition state links two permutational isomers of the Mackay[99] icosahedron via a concerted sextuple diamond-square-diamond[100] rearrangement.[101] Hence these experiments test our ability to stabilise a structure that corresponds to a saddle point for the original potential, and probe the conditions required to construct a doubly addressable system.

A. Results for LJ_5^*

If permutation-inversion isomers are not distinguished then the LJ_5 cluster has a single minimum and two distinct transition states. The transition states link permutational isomers of the minimum via degenerate rearrangements.[102] The number of permutation-inversion minima corresponding to distinct wells on the potential energy surface is $2N!/o_\alpha$ for N identical atoms and a structure with point group order o_α.[9,102–106] Hence there are $2 \times 5!/12 = 20$ versions of the D_{3h} minimum and 30 and 60 distinct versions of the two transition states (Table I).

The simple choice of addressable potential employed in the present work, with

just two distinct well depths, does not resolve all the degeneracies in energy. Instead, the 20 minima split into sets of 2, 6 and 12, as shown in the disconnectivity graphs in Figure 1. The splitting between the two larger sets varies from 2.8×10^{-5} for $\alpha = 0.75$ to 2.3×10^{-4} for $\alpha = 0.25$, values that are too small to discern in the graphs. Similarly, the transition states split into three sets of 6, four sets of 12, and one set of 24, preserving the total of 90. The remaining permutational degeneracy could be lifted by adding a random component to the pair well depths, but this step does not seem helpful in the present study.

The results in Figure 1 illustrate two general points. First, the potential defined in equation (1) has inversion symmetry, and this two-fold degeneracy is evident in the disconnectivity graphs in the left panels of the figure. Second, the landscape becomes increasingly biased towards the reference structure as α decreases, realising the design principles on which this general addressable form is based. The trend is clearer when inversion isomers are lumped together, as shown in the right hand panels of Figure 1, and this representation is retained for all the graphs that follow.

B. Results for LJ_6^*

For LJ_6 there are 30 distinct permutation-inversion isomers of the octahedral global minimum, and 360 versions of the C_{2v} local minimum. There are three different transition state structures, with 720 versions for two of them and 360 versions for the third. For $\alpha = 0.75$ and 0.5 the octahedral minima split into two sets with point group order 48, twelve of order 8, and sixteen of order 6, to preserve the total of 30. The two minima with O_h symmetry correspond to

the permutational isomer chosen as the reference and its mirror image. The sets of twelve and sixteen minima have atoms exchanged in one of the twelve edges of the octahedron, or three atoms permuted in a face. There are two possible permutations for each of eight faces, producing sixteen distinguishable minima. The splittings that occur for lower symmetry minima and transition states can also be analysed in this way, but details are omitted for brevity.

Table I summarises the results for the smaller clusters as a function of α. Here it is noteworthy that the number of stationary points can change when $\alpha \neq 1$, and an isomorphism with the unperturbed system need not be preserved. The disconnectivity graphs for LJ_6^* in Figure 2 lump inversion isomers together, and illustrate how the complexity of the landscape increases combinatorially with system size. The addressable potential produces a well-defined global minimum as α decreases.

C. Results for LJ_7^*

For LJ_7 the D_{5h} global minimum (pentagonal bipyramid) has 504 distinct permutation-inversion isomers. Two minima with point group order 6 support 1680 versions each, with 5040 for the remaining C_2 minimum, giving a total of 8904. For the twelve distinct transition states identified, there is one with point group D_{3d}, three with point group C_{2v}, five with point group C_s, and three with point group C_1. The total number of permutation-inversion isomers is therefore $840 + 3 \times 2520 + 5 \times 5040 + 3 \times 10080 = 63840$.

Even when inversion isomers are lumped together, the details of the corresponding disconnectivity graphs are difficult to distinguish (Figure 3), although

the emergence of an addressable global minimum for smaller α is still clear. The resulting landscape has the form that we associate with efficient self-organisation. The 'palm tree' structure[91,102,107] corresponds to a well-defined global potential energy minimum with no other low-lying structures separated by high barriers, which would constitute kinetic traps. The resulting free energy minimum should therefore be kinetically accessible over a wide range of temperature,[12,108] producing a landscape with 'funnelling' properties[109–111] in terms of convergent kinetic pathways.[112] The landscape entropy,[113–115] defined in terms of the energy density of local minima, decreases systematically on downhill paths to the global minimum.

Further coarse-graining of the landscape may help to visualise the above properties more clearly. Minima with no direct connections to a lower energy minimum lie at the terminus of a monotonic sequence.[116–119] Hence it is not possible to reach a lower minimum via a rearrangement corresponding to a single transition state. The disconnectivity graph including only these minima provides a simplified view of the landscape that preserves the structure and connectivity of the monotonic sequence basins. An example is shown in Figure 4 for LJ_7^* with $\alpha = 0.25$. The 14 minima in the resulting graph are all permutational isomers of the global minimum pentagonal bipyramid. The next seven minima above the global minimum correspond to a pair exchange of adjacent equatorial atoms, or exchange of an axial and an equatorial atom, defining a set of five and a set of two degenerate configurations, with very similar energies. In each case two nearest-neighbour contacts in the reference structure are lost, resulting in an energy increase of approximately 2α (in units of ϵ). The next set of five minima lies about 3α above the global minimum, with three nearest-neighbour contacts lost within

the equatorial set of atoms. All five contacts in the equator are lost in the highest minimum, and this isomer is unique aside from the corresponding permutation-inversion isomer. In each case the permutation-inversion isomers have again been lumped together.

The structures corresponding to monotonic sequence termini are actually a form of *biminimum*, as defined in recent work on global optimisation for multicomponent systems.[120,121] For nanoalloy clusters, a biminimum is defined as a local minimum whose energy cannot be lowered by interchanging any inequivalent particles and reminimising. We can generalise this idea to *multiminima*,[121,122] where the energy cannot be lowered further by perturbations in any number of different metric spaces, including continuous coordinate space, identity swaps for different atoms, or exchanges in some other category. Hence the monotonic sequence termini qualify as biminima because the energy cannot be lowered by stepping to any adjacent minimum that is connected by a single transition state. This viewpoint can be very helpful for interpreting the disconnectivity graphs that retain only the corresponding subset of local minima, which is a convenient way to provide a visualisation of large databases. The underlying structure that is revealed can provide important insight into the emergence of observable properties, such as multiple relaxation time scales and features in the heat capacity.[92,93,123–128]

D. Results for LJ^*_{13}

For LJ^*_{13} we choose to present the landscape for $r_c = 1.2$ and $\alpha = 0.25$, which exhibits a well-defined self-organising structure. In contrast to the smaller clusters, above, we can sample only a small fraction of the stationary points for this

system, and the focus is on converging the database in the low energy range of interest. To examine global properties, such as the heat capacity, at higher temperatures would require additional calculations to provide a proper representation of higher energy minima corresponding to the liquid-like phase. We plan to employ basin-sampling[115] for this purpose in future work.

Two disconnectivity graph representations of the potential energy landscape are compared in Figure 5, and characteristics of the connected component of the database that includes the global minimum are summarised in Table II. The panel on the left includes only minima in the lowest 2000 that correspond to monotonic sequence termini, while the graph on the right includes the lowest 500 minima. Some of the latter structures do not correspond to the biminima defined by monotonic sequence termini, because they are directly connected to a lower energy structure, so they do not appear in the left panel. The two graphs are complementary: for example, the second-lowest minima are not present in the left panel because they are connected to the global minimum. However, the structure of this part of the landscape is clear, because these minima correspond to exchange of two atoms in each of the thirty edges of the icosahedron. Both representations correspond to the organisation we would associate with efficient relaxation to the global minimum, again suggesting a successful addressable design.

E. Results for $LJ_{55}^{I_h}$, $LJ_{55}^{O_h}$ and $LJ_{55}^{I_h+O_h}$

For this larger cluster two additional tests were conduced to investigate the addressability of a target structure corresponding to a higher energy stationary point (in this case a transition state), and a doubly addressable potential that targets

two reference geometries. The capability to select different morphologies, and to design multifunnel landscapes, could play a particularly important role in future design of *multifunctional* landscapes[129] and molecular switches, such as rings and cages that might open and close in response to external conditions.[63]

The double reference potential defined in equation (3) can be made more flexible using different cutoffs and well depths for the two references. To locate suitable parameters the pathway mediated by the O_h transition state between two I_h minima was analysed systematically, first to stabilise the transition state as a minimum, and then to adjust the two target structures to have comparable potential energy. Using cutoffs of 1.2 for both references produces a range of α values where both targets correspond to minima, and choosing $\alpha = 0.2$ gives a pathway between these minima mediated by a single transition state of lower symmetry, with similar barriers in the two directions. Hence the present investigation did not require additional flexibility beyond the form of equation (3).

For each system, $\mathrm{LJ}_{55}^{I_h}$, $\mathrm{LJ}_{55}^{O_h}$ and $\mathrm{LJ}_{55}^{I_h+O_h}$, stationary point databases were expanded until the appearance of the disconnectivity graphs in the low energy range of interest ceased to change significantly. The number of minima and transition states in the connected component of each database containing the global minimum is given in Table II, and the disconnectivity graphs retaining monotonic sequence termini that lie in the lowest 2000 minima are shown in Figure 6. In each case the second-lowest minimum is added to this set for comparison.

These results show that the simple addressable potentials defined in §I can indeed selectively stabilise different morphologies, which need not be local minima for the original potential. Furthermore, the framework can be extended to multiple targets in a straightforward way. Comparison of the low-lying regions of the po-

tential energy landscapes for the three systems in Figure 6 suggests some further trends, which need to be tested in future work to determine if they are generally applicable. Specifically, the addressable landscape is simplest when the target structure is a permutational isomer of the global potential minimum for $LJ_{55}^{I_h}$. A minimum derived from the icosahedron is identified as the second-lowest minimum for $LJ_{55}^{O_h}$ (it does not lie at the end of a monotonic sequence), and this landscape exhibits more low-lying minima connected to the O_h minimum by larger, but still relatively small, barriers. The structure of the doubly-addressable landscape appears intermediate in character between the graphs obtained for $LJ_{55}^{I_h}$ and $LJ_{55}^{O_h}$.

IV. GLOBAL PROPERTIES OF ADDRESSABLE CLUSTERS

Two properties were calculated to examine the emergence of characteristics that reflect global features of the underlying potential energy landscape, namely the heat capacity and a recently introduced measure of frustration.[130] Here we consider the smaller clusters, where the stationary point databases are probably complete. The heat capacity was calculated using the superposition approach, where the partition function is written as a sum over contributions from the local minima.[102,131–135] The harmonic approximation was employed for the local vibrational densities of states, corresponding to standard normal mode analysis. Hence we neglect well anharmonicity, but include the landscape anharmonicity, which derives from the distribution of local minima.[113–115] Anharmonicity arising from the deviation of local potential wells from the normal mode Hamiltonian generally introduces systematic shifts in heat capacity peaks, since higher energy minima

corresponding to a liquid-like phase usually have lower vibrational frequencies. However, it is the effect of landscape entropy that is of interest here, especially in comparing results for changes in the fractional well depth parameter α. The partition functions employed in the present work must omit symmetry numbers corresponding to the molecular point group, since the permutational isomers are counted explicitly. The addressable potential lifts some of the usual permutational degeneracy, even when some geometrical symmetry elements are retained in certain stationary points.

Competition between low energy structures with different morphologies separated by high barriers has been associated with 'frustration'.[109,136] To analyse the degree of frustration in the landscapes of these addressable clusters we calculated

$$f(T) = \sum_{\gamma \neq \mathrm{gmin}} p_\gamma^{\mathrm{eq}}(T) \left(\frac{V_\gamma^\dagger - V_{\mathrm{gmin}}}{V_\gamma - V_{\mathrm{gmin}}} \right), \tag{6}$$

where T is the temperature, $p_\gamma^{\mathrm{eq}}(T)$ is the equilibrium occupation probability of minimum γ, V_{gmin} signifies the potential energy of the lowest minimum in the database, $V_\gamma^\dagger$ is the potential energy of the highest transition state on the lowest energy path between minimum γ and this lowest minimum, and V_γ is the potential energy of minimum γ. $f(T)$ employs barrier thresholds $V_\gamma^\dagger$ in the same way as a scheme that we introduced in earlier work to help refine databases produced by discrete path sampling to remove artificial frustration.[137] This approach was, in turn, derived from an analogous index based on free energies that we used earlier,[138] which extends measures based on stability[139] and energy gaps[140,141] by including explicit barrier information.

$p_\gamma^{\mathrm{eq}}(T)$ was calculated using harmonic vibrational densities of states, and $V_\gamma^\dagger$ is the potential energy of the lowest transition state for which a path exists between

minimum γ and the global minimum. The most insightful index is probably $\widetilde{f}$ calculated from the renormalised probabilities $\widetilde{p}_\gamma^{\rm eq} = p_\gamma^{\rm eq}/(1-p_{\rm gmin}^{\rm eq})$. This quantity reflects the renormalised relative populations of the minima when the temperature dependence of the global minimum is removed. Here $p_{\rm gmin}^{\rm eq}$ is summed over degenerate global minima, i.e. all permutation-inversion isomers for $\alpha = 1$, and the two enantiomers for $\alpha < 1$.

To calculate $\widetilde{f}$ we require databases where the connectivity of the local minima is defined by transition states. We obtained $V_\gamma^\dagger$ via the superbasin analysis[90] that yields the disconnectivity graphs, identifying the energy threshold below which the lowest minimum is no longer accessible from minimum γ. Results for $\widetilde{f}(T)$, which we will refer to as the *frustration index*, are shown in Figure 7.

The heat capacity peaks shift to higher temperature as α varies through 0.75, 0.5 to 0.25. This trend reflects the increasing energy gap between the global minimum and the higher energy minima, as expected for the finite system analogue of a first order phase transition. The global minimum is stabilised by potential energy, and the other minima can be viewed collectively as a higher entropy phase-like form. Here the variation in potential energy with α is more significant than changes in the vibrational entropy, consistent with results for clusters containing a dopant atom.[122] The results for the unperturbed potential can be qualitatively different because of the higher degeneracies. For example, C_v for LJ_5 is temperature independent, because all the minima are degenerate.

The trends for the *frustration index* are related to those for C_v. For LJ_5^* the minima split into two sets for $\alpha < 1$, and so there is no temperature dependence once the effect of the global minimum is removed. For the other two sizes, $\widetilde{f}(T)$ usually decreases through the α values 0.75, 0.5 to 0.25, reflecting a decrease

in frustration. This observation is consistent with the landscape becoming more biased (funnelled) towards the global minimum. Again, the unperturbed potential behaves differently, because of the unresolved permutational degeneracies, which result in a qualitatively different energy level spectrum.

Hence we see that both C_v, a key experimental observable, and $\widetilde{f}(T)$, which is readily obtained from the stationary point databases, can provide insight into the global structure of the landscape in a compact form. Experimentally, the yield of a target structure will depend upon the starting conditions, annealing schedule, and the observation time scale. An upper bound on the yield in the limit of slow enough annealing can be obtained from the equilibrium occupation probability of the global minimum, which is used in the renormalisation of $\widetilde{f}(T)$. This probability is plotted as a function of temperature and α for LJ_5^*, LJ_6^*, and LJ_7^* in Figure 8, again using the approximation of harmonic vibrational densities of states. There is a clear trend for $\alpha \neq 1$, with larger equilibrium yields of the target as α decreases.

V. MULTIPLE TARGET CLUSTERS

The energy landscapes for selected sizes between two and 120 copies of LJ_5* were surveyed using successive basin-hopping global optimisation. Initial runs were conducted using the local rigid body formulation[142] to sample the configuration space of clusters with particles maintained in their most favourable addressable sites for the target cluster. A range of values for $\epsilon^{\rm rep}$ between 500 and 15000 was considered. The lowest five or ten minima obtained in each case with local rigidification were then fully relaxed with all the $\epsilon^{\rm rep}$ values considered. The

lowest minimum for each $\epsilon^{\rm rep}$ was then used to initiate a new global optimisation run, and this procedure was iterated until the lowest fully relaxed structure in each case was unchanged.

The most systematic tests were conducted for 15 copies of a five-particle addressable cluster with $\alpha = 0.1$. For $500 < \epsilon^{\rm rep} < 3000$ interesting structures, including closed shells and incomplete shells, are the lowest minima for clusters of rigid LJ_5* molecules with all the particles in their most favourable positions in each local rigid body. However, these structures change significantly on unconstrained relaxation, and below $\epsilon^{\rm rep}$ of order 3000 the individual 5-particle clusters distort significantly. Above $\epsilon^{\rm rep}$ around 4000 the cluster structure is preserved on relaxation, and the lowest minima located on basin-hopping both with and without local rigid body constraints correspond to stacked rings of eight plus seven or nine plus six trigonal bipyramids (Figure 9). For $\epsilon^{\rm rep} = 8000$ and above the lowest minimum located is a 15-membered ring. Other favourable morphologies include combinations of rings and chains. The existence of low-lying minima corresponding to rotaxane structures, with a chain passing through a ring, is particularly interesting (Figure 9). These motifs are also found when more copies of the target are included; some examples are shown in Figure 10, including a catenane formed from two 15-cluster rings. The lowest minima located in short basin-hopping runs for 60 and 120 copies of the target involve convoluted chains. All of these structures were fully relaxed with no constraints. Locating the true global minima will require much longer runs, but the structures identified so far strongly suggest that all the favourable aggregates will contain arrays of rings and chains. The possibility of exploiting hierarchical design to produce complex morphologies, including knots, clearly warrants further investigation. To provide some indication of the

likely complexity encoded by such aggregates we conclude this preliminary survey with a detailed analysis of the landscape for the LJ_5* dimer.

Two copies of LJ_5* were first considered, with each one treated as a local rigid body based on the monomer global minimum. The number of distinct local minima on this simplified landscape falls from six for $\epsilon^{\rm rep} = 500$, to four for $\epsilon^{\rm rep} = 1000$, two for $\epsilon^{\rm rep} = 3000$, and then a single minimum for $\epsilon^{\rm rep} = 6000$ and above. Here we have not counted minima with positive energies, where the two rigidified LJ_5* clusters interpenetrate. A detailed analysis of the unconstrained landscape was then performed for $\epsilon^{\rm rep} = 8000$ using the discrete path sampling tools encoded in our OPTIM[143] and PATHSAMPLE[94] programs.

The disconnectivity graph where all permutational isomers are distinguished, with enantiomers lumped together, is shown in Figure 11. Rearrangements between the 16 subfunnels correspond to exchange of particles with the same address. There are $2^5 = 32$ permutations for every minimum, and the 16 funnels each contain two permutational isomers of the global minimum along with the corresponding enantiomers. These two permutational isomers are related by exchange of all pairs of particles with the same address.

The hierarchical organisation of the landscape is clear, with the 16 funnels corresponding to subsets of local minima that are connected in groups of four via a barrier of around $2.88984\,\epsilon$; the lowest minima within different subfunnels in the latter groups can interconvert via pathways mediated by two transition states. Both permutational isomers of the global minimum are illustrated for each funnel in Figure 11, with the pairs of equivalent atoms coloured bright and pale shades of blue, red, orange, yellow and silver in a consistent numbering scheme. The database that underlies this graph contains 29620 minima and 374474 transition

states. To check the convergence of the sampling, symmetry was not exploited, and the resulting hierarchical structure emerged systematically as new connections between minima were located.

A magnification of the low-energy region of one sub-funnel is shown in Figure 12. The graph is based on the lowest 171 minima in this region, which includes all the structures with distinguishable correctly addressed target clusters, along with some higher energy configurations corresponding to wrongly addressed components; there is a clear energetic separation between these sets. A few minima exhibiting small geometrical perturbations of one of the two trigonal bipyramids also appear, with comparable energies to structures with wrongly addressed components. One of these is illustrated in Figure 12 (the penultimate structure on the right of the disconnectivity graph).

The transition state that permits interchange between all permutational isomers of the global minimum corresponds to an overall barrier of $3.44322\,\epsilon$. Rearrangements within each subfunnel correspond to changes in address for the particles within one or both of the distinct LJ_5* clusters, along with possible geometrical perturbations of the target trigonal bibyramid. The overall yield of correctly addressed targets will depend on the energy gap between these targets and the lowest alternative minima. The relatively large gap observed here means that dimers composed of the correct targets are thermodynamically favourable and kinetically accessible over a significant range of temperature (or microcanonical total energy).

VI. CONCLUSIONS

The present results show that altering the strength of interparticle interactions to favour contacts that are present in a reference structure can provide a powerful design tool for addressable structures. Similar constructions have been used before to simplify analysis of biomolecules, and a generalisation to clusters and soft and condensed matter is straightforward. The present approach could be used for pairwise and many-body potentials, and the biasing could be tuned more extensively using a range of well depth and cutoff parameters.

In fact, a simple implementation with a single cutoff and a uniform reduction in the strength of non-nearest-neighbour interactions, is sufficient to produce addressable clusters for the systems considered here. Landscapes with increasing self-organising characteristics are obtained as the bias changes systematically. It is also possible to define target structures corresponding to a relatively high energy transition state with a completely different morphology in the 55-atom cluster, and to design a doubly-addressable system for this geometry and a permutational isomer of the global minimum.

For multiple copies of the target cluster, introducing a repulsive term between particles with the same address can produce aggregates with distinct copies of the target. As the repulsive parameter increases, the favourable morphologies for these aggregates include rings and chains, and complex topologies appear, such as rotaxanes and catenanes. Realising these exotic structures experimentally will require a mapping from the model potential, in terms of competing energy and length scales, to details of the intermolecular potential for the building blocks involved. We have previously exploited such principles to suggest designs for

helical assemblies.[129,144,145]

Visualising the underlying potential energy landscape using disconnectivity graphs provides insight into the evolution with increasing bias. Global thermodynamic properties are reflected in the heat capacity, while the frustration index includes information about barrier heights, and hence kinetic accessibility. Identifying structures that are only directly connected to higher-lying minima as biminima, provides a convenient way to simplify the disconnectivity graphs for larger databases. These tools could play an important role in future design efforts for addressable systems, including multifunctional properties that we associate with multifunnel landscapes.[146]

Acknowledgements

I am very grateful to Brooke Husic, Dmitri Schebarchov and Aleks Reinhardt for comments on the first draft of this manuscript, to John Morgan for his assistance with some of the graphics, and to Prof. Robert Jack for insight into previous work on glassy systems.

REFERENCES

[1]Y. Ke, L. L. Ong, W. M. Shih and P. Yin, Science **338**, 1177 (2012).

[2]A. Reinhardt and D. Frenkel, Phys. Rev. Lett. **112**, 238103 (2014).

[3]W. M. Jacobs, A. Reinhardt and D. Frenkel, Proc. Natl. Acad. Sci. USA **112**, 6313 (2015).

[4]G. Meng, N. Arkus, M. P. Brenner and V. N. Manoharan, Science **327**, 560 (2010).

[5]J. C. Crocker, Science **327**, 535 (2010).

[6]D. J. Wales, ChemPhysChem **11**, 2491 (2010).

[7]S. Hormoz and M. P. Brenner, Proc. Nat. Acad. Sci. USA **108**, 5193 (2011).

[8]D. Frenkel and D. J. Wales, Nature Materials **10**, 410 (2011).

[9]F. Calvo, J. P. K. Doye and D. J. Wales, Nanoscale **4**, 1085 (2012).

[10]J. W. Collins, *Self-assembly of colloidal spheres with specific interactions. doctoral dissertation, harvard university. http://nrs.harvard.edu/urn-3:hul.instrepos:12274201* (2014).

[11]B. Derrida, Phys. Rev. B **24**, 2613 (1981).

[12]J. D. Bryngelson and P. G. Wolynes, Proc. Natl. Acad. Sci. USA **84**, 7524 (1987).

[13]J. E. Jones and A. E. Ingham, Proc. R. Soc. A **107**, 636 (1925).

[14]Y. Ueda, H. Taketomi and N. Gō, Biopolymers **17**, 1531 (1978).

[15]S. Franz and G. Parisi, Phys. Rev. Lett. **79**, 2486 (Sep 1997).

[16]L. Berthier and R. L. Jack, Phys. Rev. Lett. **114**, 205701 (May 2015).

[17]R. L. Jack and J. P. Garrahan, Phys. Rev. Lett. **116**, 055702 (Feb 2016).

[18]C. J. Fullerton and R. L. Jack, Phys. Rev. Lett. **112**, 255701 (2014).

[19]R. L. Jack and C. J. Fullerton, Phys. Rev. E **88**, 042304 (2013).

[20]L. Berthier and W. Kob, Phys. Rev. E **85**, 011102 (2012).

[21]W. Kob and L. Berthier, Phys. Rev. Lett. **110**, 245702 (2013).

[22]S. Karmakar and G. Parisi, Proc. Natl Acad. Sci. USA **110**, 2752 (2013).

[23]C. Cammarota and G. Biroli, ArXiv e-prints (2011).

[24]C. Cammarota and G. Biroli, Proc. The National Acad. Sci. The United States Am. **109**, 8850 (2012).

[25]C. Cammarota and G. Biroli, J. Chem. Phys. **138**, 12A547 (2013).

[26]K. Kim, Europhysics Lett. **61**, 790 (2003).

[27]M. Ozawa, W. Kob, A. Ikeda and K. Miyazaki, PNAS **112**, 6914 (2015).

[28]C. Cammarota, G. Gradenigo and G. Biroli, Phys. Rev. Lett. **111**, 107801 (2013).

[29]Y. W. Li, W. S. Xu and Z. Y. Sun, J. Chem. Phys. **140**, 124502 (2014).

[30]A. Cavagna, T. S. Grigera and P. Verrocchio, J. Chem. Phys. **136**, 204502 (2012).

[31]A. Cavagna, T. S. Grigera and P. Verrocchio, Phys. Rev. Lett. **98**, 187801 (2007).

[32]C. Cammarota and A. Cavagna, J. Chem. Phys. **127**, 214703 (2007).

[33]G. Biroli, J.-P. Bouchaud, A. Cavagna, T. S. Grigera and P. Verrocchio, Nature Phys. **4**, 771 (2008).

[34]C. Cammarota, A. Cavagna, G. Gradenigo, T. S. Grigera and P. Verrocchio, J. Stat. Mech. - Theory Exp. p. L12002 (2009).

[35]A. Cavagna, T. S. Grigera and P. Verrocchio (Jun 2010).

[36]L. Berthier, P. Charbonneau and S. Yaida, arxiv **1510.06320** (2015).

[37]S. Yaida, L. Berthier, P. Charbonneau and G. Tarjus, arxiv **1511.03573** (2015).

[38]M. M. Tirion, Phys. Rev. Lett. **77**, 1905 (1996).

[39]F. Tama and Y. H. Sanejouand, Prot. Eng. **14**, 1 (2001).

[40]A. R. Atilgan, S. R. Durell, R. L. Jernigan, M. C. Demeril, O. Keskin and I. Bahar, Biophys. J. **80**, 505 (2001).

[41]M. K. Kim, G. S. Chirikjan and R. L. Jernigan, J. Mol. Graph. Model **21**, 151 (2002).

[42]D. Ming, Y. Kong, M. A. Lambert, Z. Huang and J. Ma, Proc. Natl. Acad. Sci. USA **99**, 8620 (2002).

[43]M. Delarue and Y. H. Sanejouand, J. Mol. Biol. **320**, 1011 (2002).

[44]M. Ikeguchi, J. Ueno, M. Sato and A. Kidera, Phys. Rev. Lett. **94**, 078102 (2002).

[45]W. Zheng and S. Doniach, Proc. Natl. Acad. Sci. USA **100**, 13253 (2003).

[46]N. Reuter, K. Hinsen and J. J. Lacapère, Biophys. J. **85**, 2186 (2003).

[47]C. Xu, D. Tobi and I. Bahar, J. Mol. Biol. **333**, 153 (2003).

[48]F. Tama, M. Valle, J. Franks and C. L. Brooks, Proc. Natl. Acad. Sci. USA **100**, 9319 (2003).

[49]P. Maragakis and M. Karplus, J. Mol. Biol. **352**, 807 (2005).

[50]I. Bahar and A. J. Rader, Curr. Op. Struct. Biol. **15**, 586 (2005).

[51]P. C. Whitford, O. Miyashita, Y. Levy and J. N. Onuchic, J. Mol. Biology **366**, 1661 (2007).

[52]W. Zheng, B. R. Brooks and G. Hummer, Proteins **69**, 43 (2007).

[53]A. Korkut and W. A. Hendrickson, Proc. Nat. Acad. Sci. USA **106**, 15667 (2009).

[54]M. Lu and J. Ma, in *Energy Flow in Proteins*, edited by D. Leitner and J. Straub, pp. 229–245, CRC Press, Boca Raton (2009).

[55]C. Peng, L. Zhang and T. Head-Gordon, Biophysical J. **98**, 2356 (2010).

[56]P. Batista, C. Robert, J. D. Marechal, M. Hamida-Rebai, P. Pascutti, P. Bisch and D. Perahia, Phys. Chem. Chem. Phys. **12**, 2850 (2010).

[57]R. B. Best, Y.-G. Chen and G. Hummer, Structure **13**, 1755 (2005).

[58]K.-i. Okazaki, N. Koga, S. Takada, J. N. Onuchic and P. G. Wolynes, Proc. Nat. Acad. Sci. USA **103**, 11844 (2006).

[59]K.-i. Okazaki and S. Takada, Proc. Nat. Acad. Sci. USA **105**, 11182 (2008).

[60]Q. Lu and J. Wang, J. Am. Chem. Soc. **130**, 4772 (2008).

[61]Z.-Z. Lai, Q. Lu and J. Wang, J. Phys. Chem. B **115**, 4147 (2011).

[62]A. Murugan, Z. Zeravcic, M. P. Brenner and S. Leibler, Proc. Natl Acad. Sci. USA **112**, 54 (2015).

[63]D. Morphew and D. Chakrabarti, Nanoscale **7**, 8343 (2015).

[64]M.-P. Valignat, O. Theodoly, J. C. Crocker, W. B. Russel and P. M. Chaikin, Proc. Natl Acad. Sci. USA **102**, 4225 (2005).

[65]P. L. Biancaniello, A. J. Kim and J. C. Crocker, Phys. Rev. Lett. **94**, 058302 (2005).

[66]A. J. Kim, P. L. Biancaniello and J. C. Crocker, Langmuir **22**, 1991 (2006).

[67]C. A. Mirkin, R. L. Letsinger, R. C. Mucic and J. J. Storhoff, Nature **382**, 607 (1996).

[68]A. P. Alivisatos, K. P. Johnsson, X. Peng, T. E. Wilson, C. J. Loweth, M. P. Bruchez and P. G. Schultz, Nature **382**, 609 (1996).

[69]S. Y. Park, A. K. R. Lytton-Jean, B. Lee, S. Weigand, G. C. Schatz and C. A. Mirkin, Nature **451**, 553 (2008).

[70]D. Nykypanchuk, M. M. Maye, D. van der Lelie and O. Gang, Nature **451**, 549 (2008).

[71]P. W. K. Rothemund, Nature **440**, 297 (2006).

[72]S. M. Douglas, H. Dietz, T. Liedl, B. Högberg, F. Graf and W. M. Shih, Nature **459**, 414 (2009).

[73]B. Wei, M. Dai and P. Yin, Nature **485**, 623 (2012).

[74]T. Wang, R. Sha, R. Dreyfus, M. E. Leunissen, C. Maass, D. J. Pine, P. M. Chaikin and N. C. Seeman, Nature **478**, 225 (2011).

[75]Y.-T. Lai, N. P. King and T. O. Yeates, Trends in Cell Biology **22**, 653 (2012).

[76]N. P. King, J. B. Bale, W. Sheffler, D. E. McNamara, S. Gonen, T. Gonen, T. O. Yeates and D. Baker, Nature **510**, 103 (2014).

[77]F. Noé and S. Fischer, Curr. Op. Struct. Biol. **18**, 154 (2008).

[78]D. Prada-Gracia, J. Gómez-Gardenes, P. Echenique and F. Fernando, PLoS Comput. Biol. **5**, e1000415 (2009).

[79]D. J. Wales, Curr. Op. Struct. Biol. **20**, 3 (2010).

[80]Z. Li and H. A. Scheraga, Proc. Natl. Acad. Sci. USA **84**, 6611 (1987).

[81]Z. Li and H. A. Scheraga, J. Mol. Struct.**179**, 333 (1988).

[82]D. J. Wales and J. P. K. Doye, J. Phys. Chem. A **101**, 5111 (1997).

[83]S. A. Trygubenko and D. J. Wales, J. Chem. Phys. **120**, 2082 (2004).

[84]G. Henkelman and H. Jónsson, J. Chem. Phys. **113**, 9978 (2000).

[85]G. Henkelman, B. P. Uberuaga and H. Jónsson, J. Chem. Phys. **113**, 9901 (2000).

[86]L. J. Munro and D. J. Wales, Phys. Rev. B **59**, 3969 (1999).

[87]G. Henkelman and H. Jónsson, J. Chem. Phys. **111**, 7010 (1999).

[88]Y. Kumeda, L. J. Munro and D. J. Wales, Chem. Phys. Lett. **341**, 185 (2001).

[89]J. N. Murrell and K. J. Laidler, Trans. Faraday Soc. **64**, 371 (1968).

[90]O. M. Becker and M. Karplus, J. Chem. Phys. **106**, 1495 (1997).

[91]D. J. Wales, M. A. Miller and T. R. Walsh, Nature **394**, 758 (1998).

[92]D. J. Wales, Mol. Phys. **100**, 3285 (2002).

[93]D. J. Wales, Mol. Phys. **102**, 891 (2004).

[94]D. J. Wales, *Pathsample: A program for generating connected stationary point databases and extracting global kinetics*, http://www-wales.ch.cam.ac.uk/software.html ().

[95]C. G. Broyden, J. Inst. Math. Appl. **6**, 76 (1970).

[96]R. Fletcher, Comput. J. **13**, 317 (1970).

[97]D. Goldfarb, Math. Comput. **24**, 23 (1970).

[98]D. F. Shanno, Math. Comput. **24**, 647 (1970).

[99]A. L. Mackay, Acta Cryst. **15**, 916 (1962).

[100]W. N. Lipscomb, Science **153**, 373 (1966).

[101]D. J. Wales, L. J. Munro and J. P. K. Doye, J. Chem. Soc. Dalton Trans. pp. 611–623 (1996).

[102]D. J. Wales, *Energy Landscapes*, Cambridge University Press, Cambridge (2003).

[103]F. G. Amar and R. S. Berry, J. Chem. Phys. **85**, 5943 (1986).

[104]M. K. Gilson and K. K. Irikura, J. Phys. Chem. B **114**, 16304 (2010).

[105]D. J. Wales and P. Salamon, Proc. Natl Acad. Sci. USA **111**, 617 (2014).

[106]M. E. Cates and V. N. Manoharan, Soft Matter **11**, 6538 (2015).

[107]D. J. Wales and T. V. Bogdan, J. Phys. Chem. B **110**, 20765 (2006).

[108]M. Karplus and A. Šali, Curr. Opin. Struct. Biol. **5**, 58 (1995).

[109]J. D. Bryngelson, J. N. Onuchic, N. D. Socci and P. G. Wolynes, Proteins: Struct., Func. and Gen. **21**, 167 (1995).

[110]J. N. Onuchic, P. G. Wolynes, Z. Luthey-Schulten and N. D. Socci, Proc. Natl. Acad. Sci. USA **92**, 3626 (1995).

[111]J. N. Onuchic, H. Nymeyer, A. E. García, J. Chahine and N. D. Socci, Adv. Prot. Chem. **53**, 87 (2000).

[112]P. E. Leopold, M. Montal and J. N. Onuchic, Proc. Natl. Acad. Sci. USA **89**, 8721 (1992).

[113]F. Sciortino, W. Kob and P. Tartaglia, J. Phys.: Condens. Matt. **12**, 6525 (2000).

[114]T. V. Bogdan, D. J. Wales and F. Calvo, J. Chem. Phys. **124**, 044102 (2006).

[115]D. J. Wales, Chem. Phys. Lett. **584**, 1 (2013).

[116]R. S. Berry and R. Breitengraser-Kunz, Phys. Rev. Lett. **74**, 3951 (1995).

[117]R. E. Kunz and R. S. Berry, J. Chem. Phys. **103**, 1904 (1995).

[118]J. P. Rose and R. S. Berry, J. Chem. Phys. **98**, 3246 (1993).

[119]K. D. Ball, R. S. Berry, R. E. Kunz, F.-Y. Li, A. Proykova and D. J. Wales, Science **271**, 963 (1996).

[120]D. Schebarchov and D. J. Wales, Phys. Rev. Lett. **113**, 156102 (2014).

[121]D. Schebarchov and D. J. Wales, Phys. Chem. Chem. Phys. **17**, 28331 (2015).

[122]B. E. Husic, D. Schebarchov and D. J. Wales, submitted for publication **000**, 0000 (2016).

[123]J. P. K. Doye, M. A. Miller and D. J. Wales, J. Chem. Phys. **110**, 6896 (1999).

[124]J. P. Neirotti, F. Calvo, D. L. Freeman and J. D. Doll, J. Chem. Phys. **112**, 10340 (2000).

[125]M. Picciani, M. Athenes, J. Kurchan and J. Tailleur, J. Chem. Phys. **135**, 034108 (2011).

[126]V. A. Sharapov and V. A. Mandelshtam, J. Phys. Chem. A **111**, 10284 (2007).

[127]F. Calvo, Phys. Rev. E **82**, 046703 (2010).

[128]D. J. Wales, Chem. Phys. Lett. **584**, 1 (2013).

[129]D. Chakrabarti and D. J. Wales, Soft Matter **7**, 2325 (2011).

[130]V. K. de Souza, J. Stevenson, J. D. Farrell, C. P. Goodrich, S. P. Niblett and D. J. Wales, submitted for publication **000**, 0000 (2016).

[131]F. H. Stillinger and T. A. Weber, Science **225**, 983 (1984).

[132]D. J. Wales, Mol. Phys. **78**, 151 (1993).

[133]F. H. Stillinger, Science **267**, 1935 (1995).

[134]B. Strodel and D. J. Wales, Chem. Phys. Lett. **466**, 105 (2008).

[135]V. A. Sharapov, D. Meluzzi and V. A. Mandelshtam, Phys. Rev. Lett. **98**, 105701 (2007).

[136]J. N. Onuchic, Z. Luthey-Schulten and P. G. Wolynes, Annu. Rev. Phys. Chem. **48**, 545 (1997).

[137]B. Strodel, C. S. Whittleston and D. J. Wales, J. Amer. Chem. Soc. **129**, 16005 (2007).

[138]J. M. Carr and D. J. Wales, J. Phys. Chem. B **112**, 8760 (2008).

[139]A. Godzik, A. Kolinski and J. Skolnick, J. Comput.-Aided Mol. Des. **7**, 397 (1993).

[140]D. U. Ferreiro, J. A. Hegler, E. A. Komives and P. G. Wolynes, Proc. Nat. Acad. Sci. USA **104**, 19819 (2007).

[141]D. U. Ferreiro, J. A. Hegler, E. A. Komives and P. G. Wolynes, Proc. Nat. Acad. Sci. USA **108**, 3499 (2011).

[142]H. Kusumaatmaja, C. S. Whittleston and D. J. Wales, J. Chem. Theor. Comput. **8**, 5159 (2012).

[143]D. J. Wales, *Optim: A program for geometry optimisation and pathway calculations*, http://www-wales.ch.cam.ac.uk/software.html ().

[144]D. Chakrabarti, S. N. Fejer and D. J. Wales, Proc. Nat. Acad. Sci. USA **106**, 20164 (2009).

[145]S. N. Fejer, D. Chakrabarti, H. Kusumaatmaja and D. J. Wales, Nanoscale **6**, 9448 (2014).

[146]Y. Chebaro, A. J. Ballard, D. Chakraborty and D. J. Wales, Sci. Rep. **5**, 10386 (2015).

cluster	α	structures		permutation-inversion isomers	
		minima	transition states	minima	transition states
LJ_5	1.00	1	2	20	90
LJ_5	0.75	3	8	20	90
LJ_5	0.50	3	8	20	90
LJ_5	0.25	3	8	20	90
LJ_6	1.00	2	3	390	1800
LJ_6	0.75	10	26	390	1896
LJ_6	0.50	10	27	390	1944
LJ_6	0.25	9	24	366	1776
LJ_7	1.00	4	12	8904	63840
LJ_7	0.75	248	1226	8904	62320
LJ_7	0.50	248	1549	8904	59418
LJ_7	0.25	230	1226	8184	47022

TABLE I. Number of distinct structures and permutation-inversion isomers corresponding to minima and transition states for addressable Lennard-Jones clusters as a function of the bias parameter α.

cluster	α	minima	transition states
LJ_{13}^{*}	0.25	396223	504668
$LJ_{55}^{I_h}$	0.20	7893	12012
$LJ_{55}^{O_h}$	0.20	19439	29915
$LJ_{55}^{I_h+O_h}$	0.20	47022	68886

TABLE II. Number of permutation isomers in the databases corresponding to minima and transition states for addressable Lennard-Jones clusters containing 13 and 55 atoms. These values include only stationary points belonging to the connected component of the database that includes the global minimum. Sampling was continued for these sizes until the low-energy region of the disconnectivity graphs appeared to have converged.

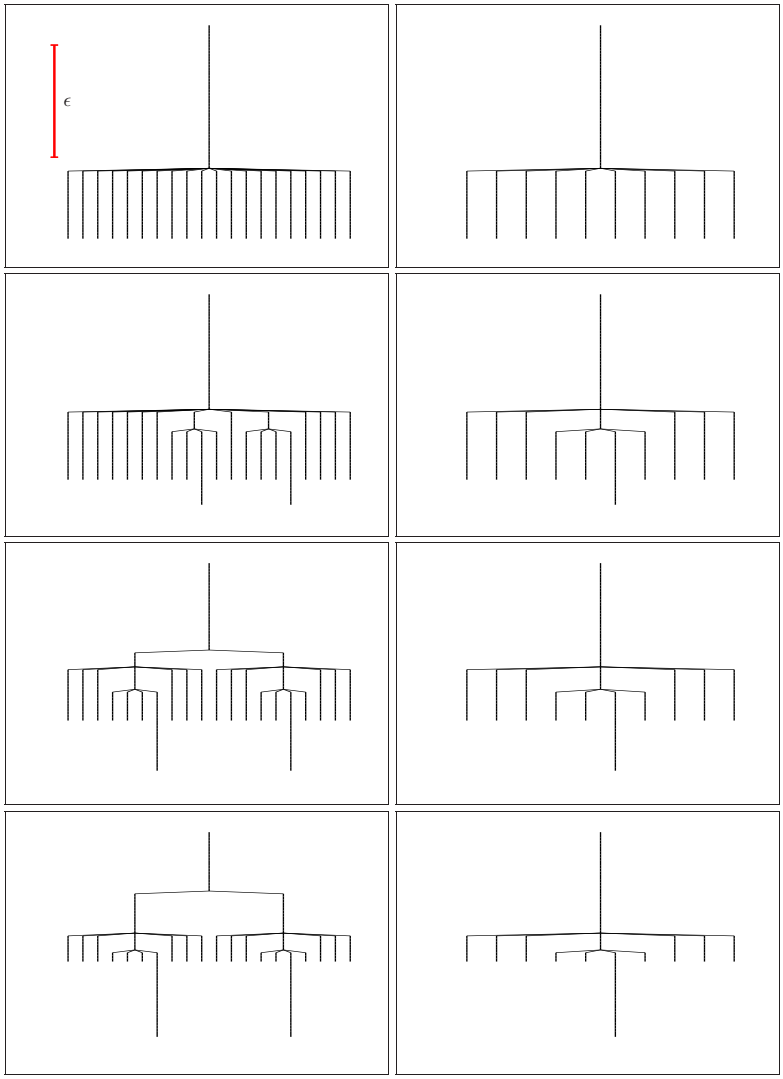

FIG. 1. Disconnectivity graphs for addressable LJ_5 for α values of 1, 0.75, 0.5, and 0.25 (top to bottom). In the left column all distinct permutation-inversion isomers are included, while in the right column pairs of inversion isomers are lumped together. The scale and energy range are the same for all panels.

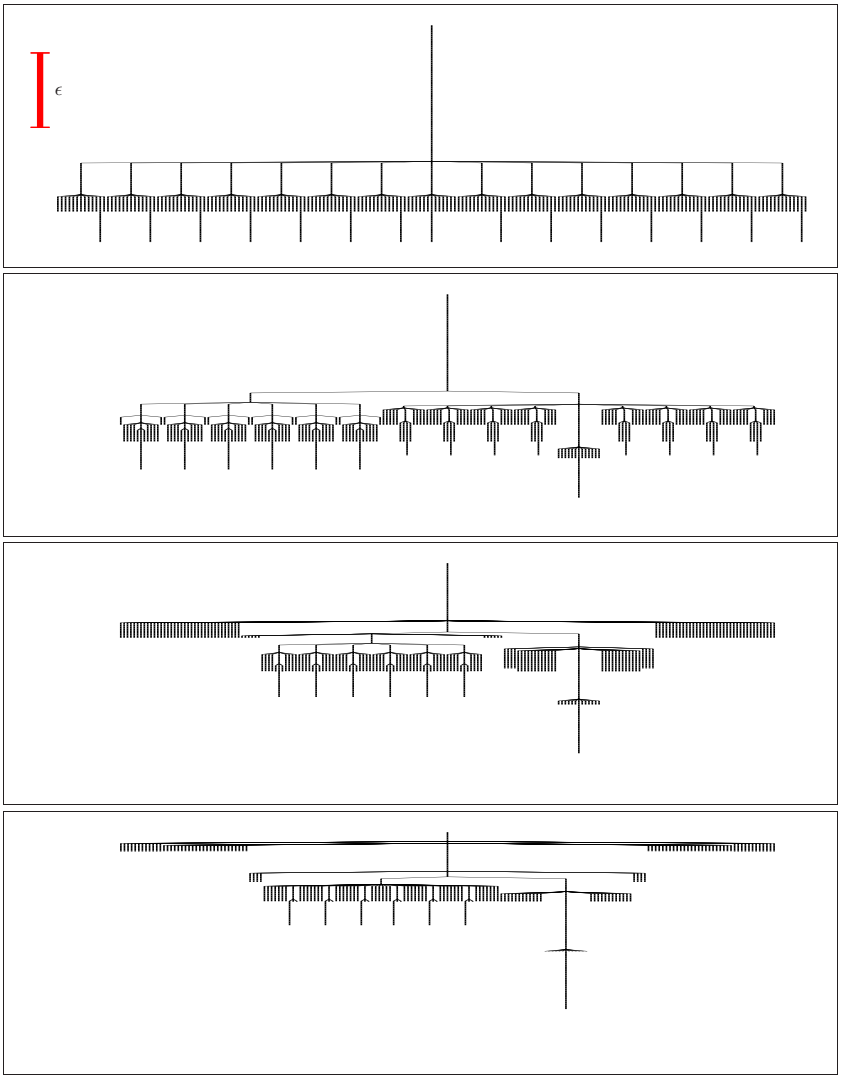

FIG. 2. Disconnectivity graphs for addressable LJ_6 for α values of 1, 0.75, 0.5, and 0.25 (top to bottom). All distinct permutational isomers are included; inversion isomers are lumped together. The scale and energy range are the same for all panels.

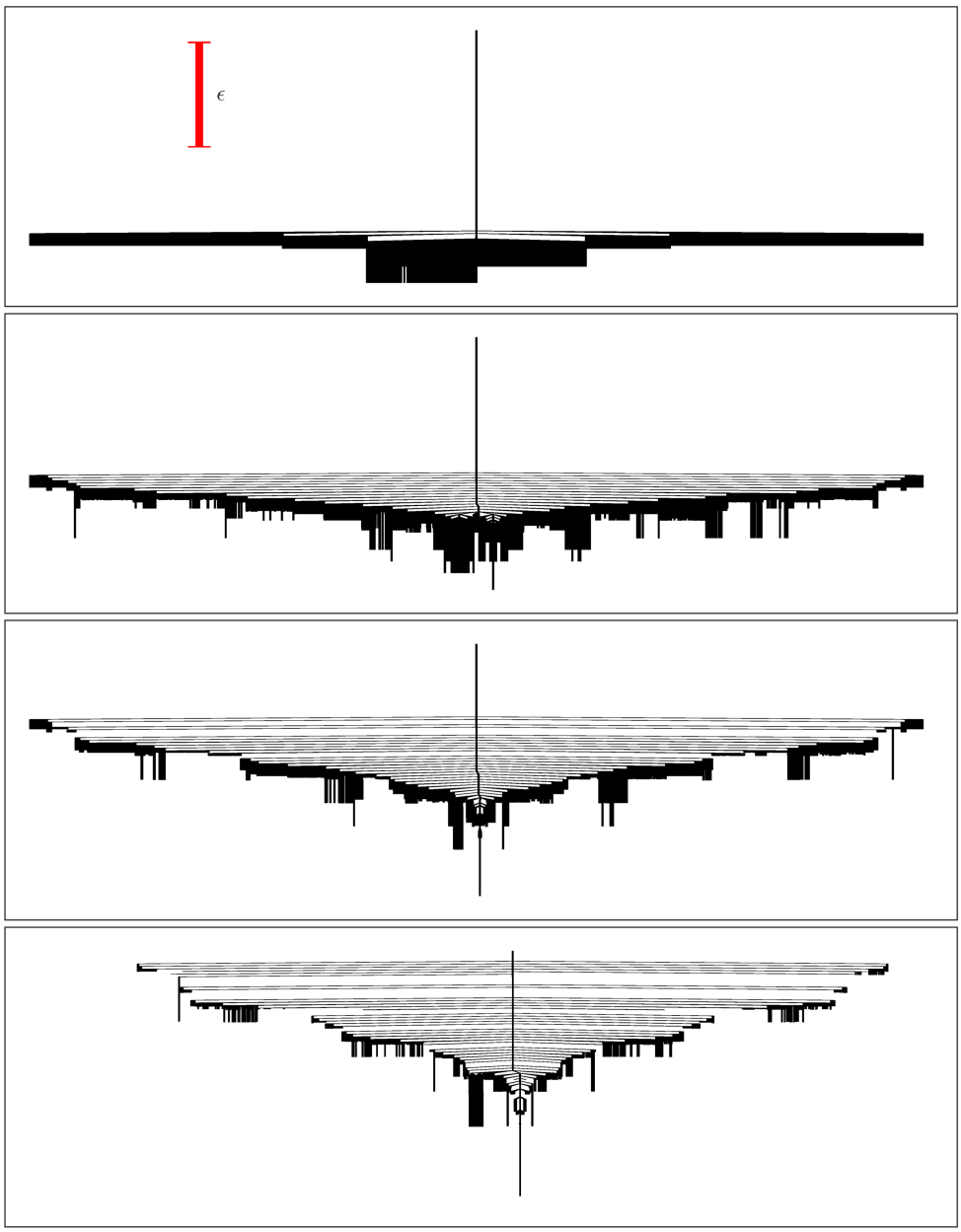

FIG. 3. Disconnectivity graphs for addressable LJ_7 for α values of 1, 0.75, 0.5, and 0.25 (top to bottom). All distinct permutational isomers are included; inversion isomers are lumped together. The scale and energy range are the same for all panels.

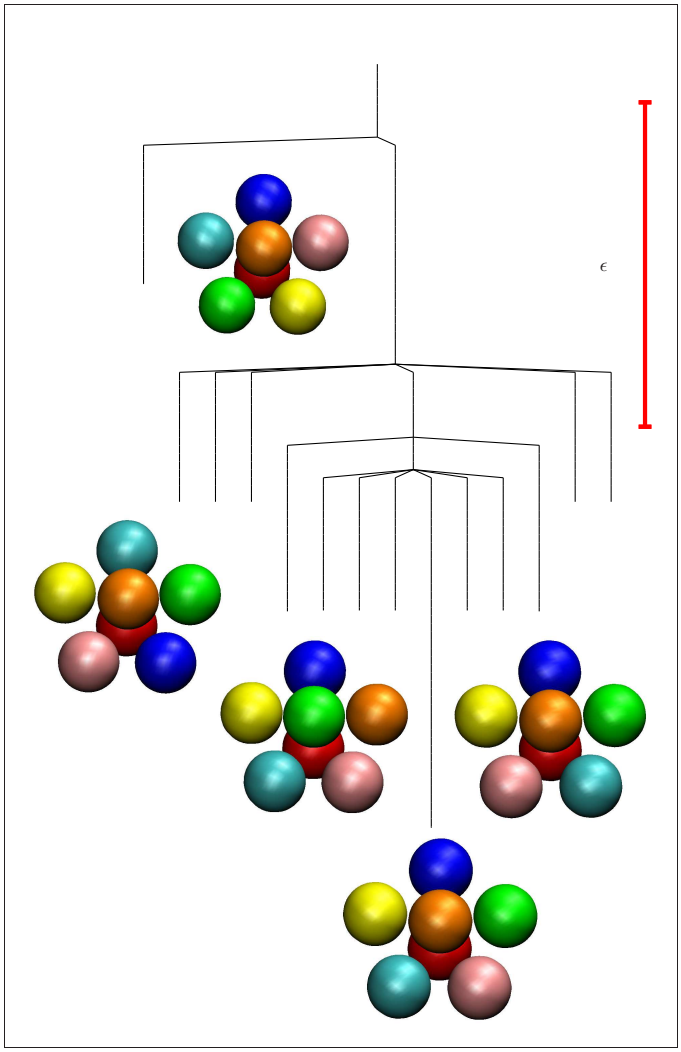

FIG. 4. Disconnectivity graph for addressable LJ_7 with $\alpha = 0.25$ showing only minima that correspond to monotonic sequence basin bottoms. The scale and energy range is the same as for the panels in Figure 3. One representative of each set of minima is shown using a consistent shading (colour online) for the atoms.

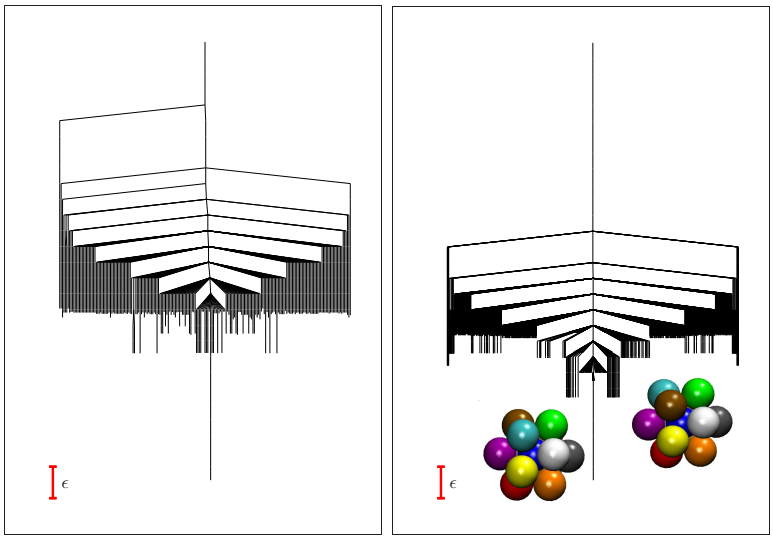

FIG. 5. Disconnectivity graphs for addressable LJ_{13} with $\alpha = 0.25$ using an isomer of the icosahedral global minimum as the reference. Inversion isomers are lumped together. Left: only minima that correspond to monotonic sequence basin bottoms in the lowest 2000 are included. Right: the lowest 500 minima, with the structures of the global minimum and one permutational isomer of the second-lowest minima illustrated. The thirty second-lowest minima have two neighbours swapped in one of the thirty edges in the outer shell.

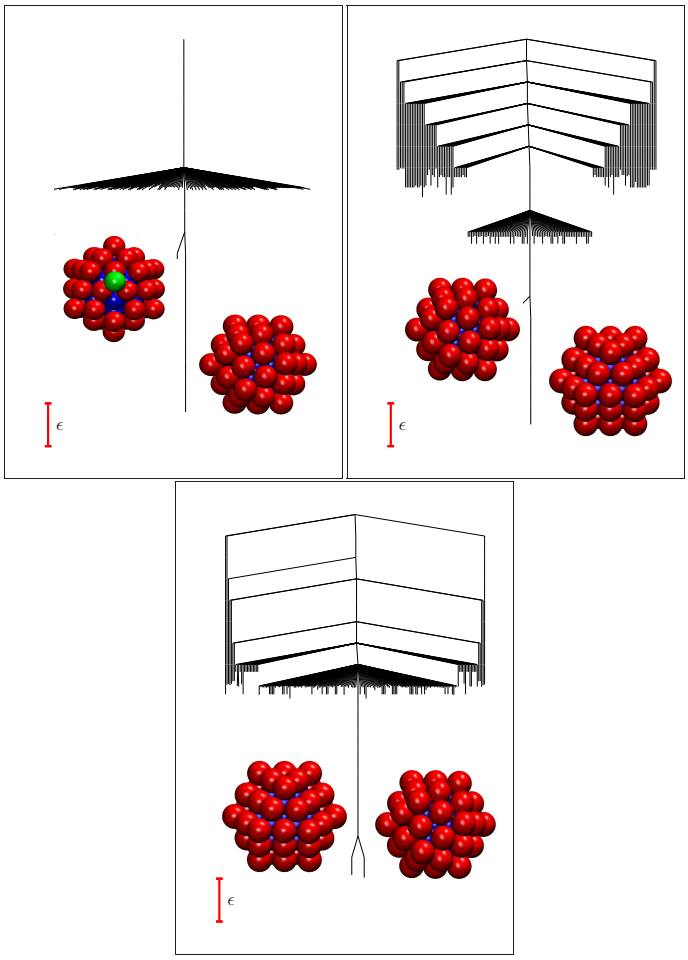

FIG. 6. Disconnectivity graphs for addressable LJ_{55} clusters with $\alpha = 0.2$ using the icosahedral global minimum as the reference (top), the cuboctahedral transition state (middle), and both stationary points (bottom). Inversion isomers are lumped together and only minima that correspond to monotonic sequence basin bottoms and also lie in the lowest 2000 are included. One branch is also included for the second-lowest minimum in each case.

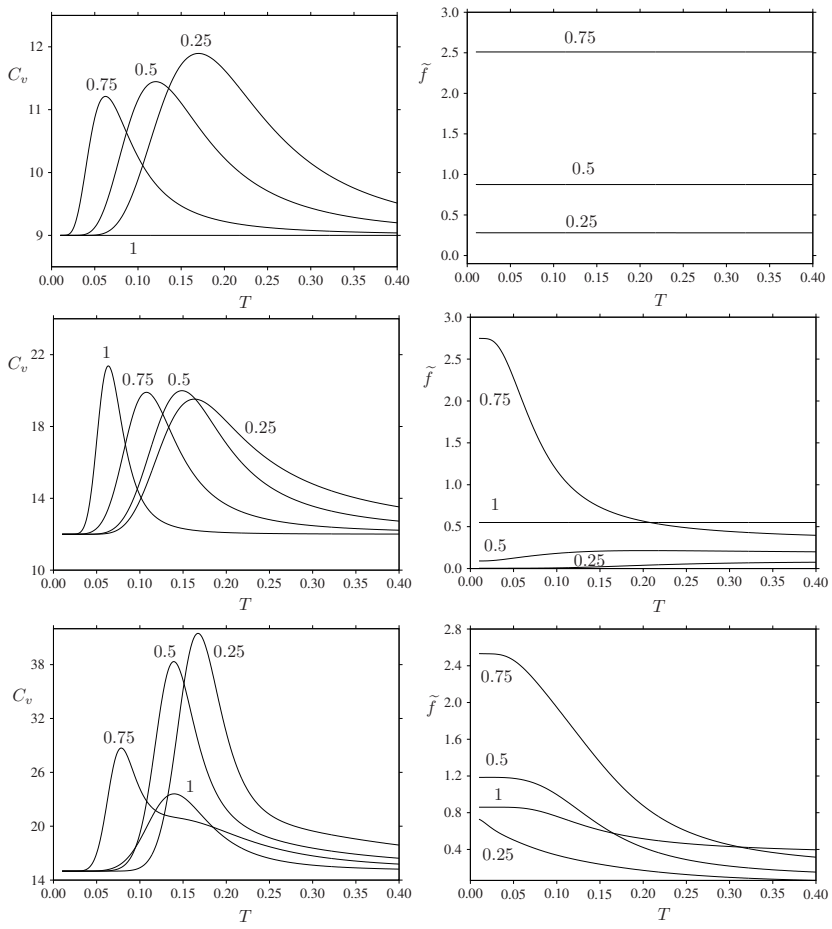

FIG. 7. Heat capacity and frustration index (left and right) as a function of temperature for addressable clusters LJ_5, LJ_6 and LJ_7 (top to bottom). In each case results are presented for α values of 1, 0.75, 0.5, and 0.25, as marked, corresponding to increasing bias towards the reference structure. C_v and T are in reduced units; $\widetilde{f}$ is dimensionless. For LJ_5 with $\alpha = 1$ all the minima are permutational isomers of the global minimum and the landscape is unfrustrated.

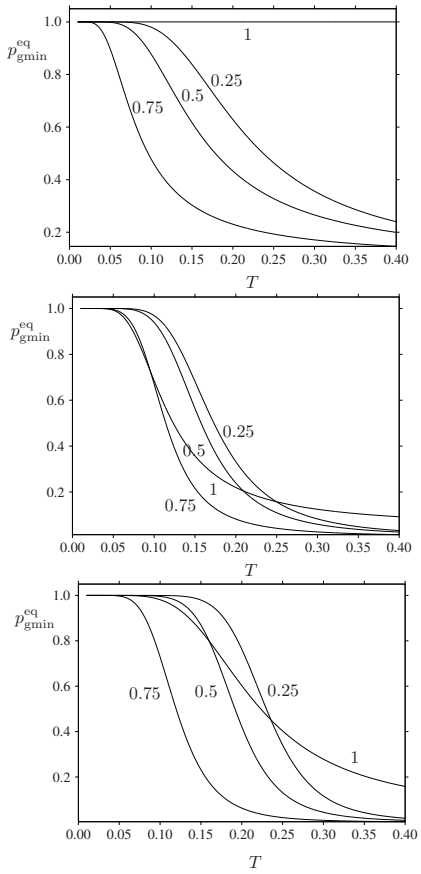

FIG. 8. Equilibrium occupation probability of the global minimum as a function of temperature for addressable clusters LJ_5, LJ_6 and LJ_7 (top to bottom). In each case results are presented for α values of 1, 0.75, 0.5, and 0.25, as marked, corresponding to increasing bias towards the reference structure. T is in reduced units.

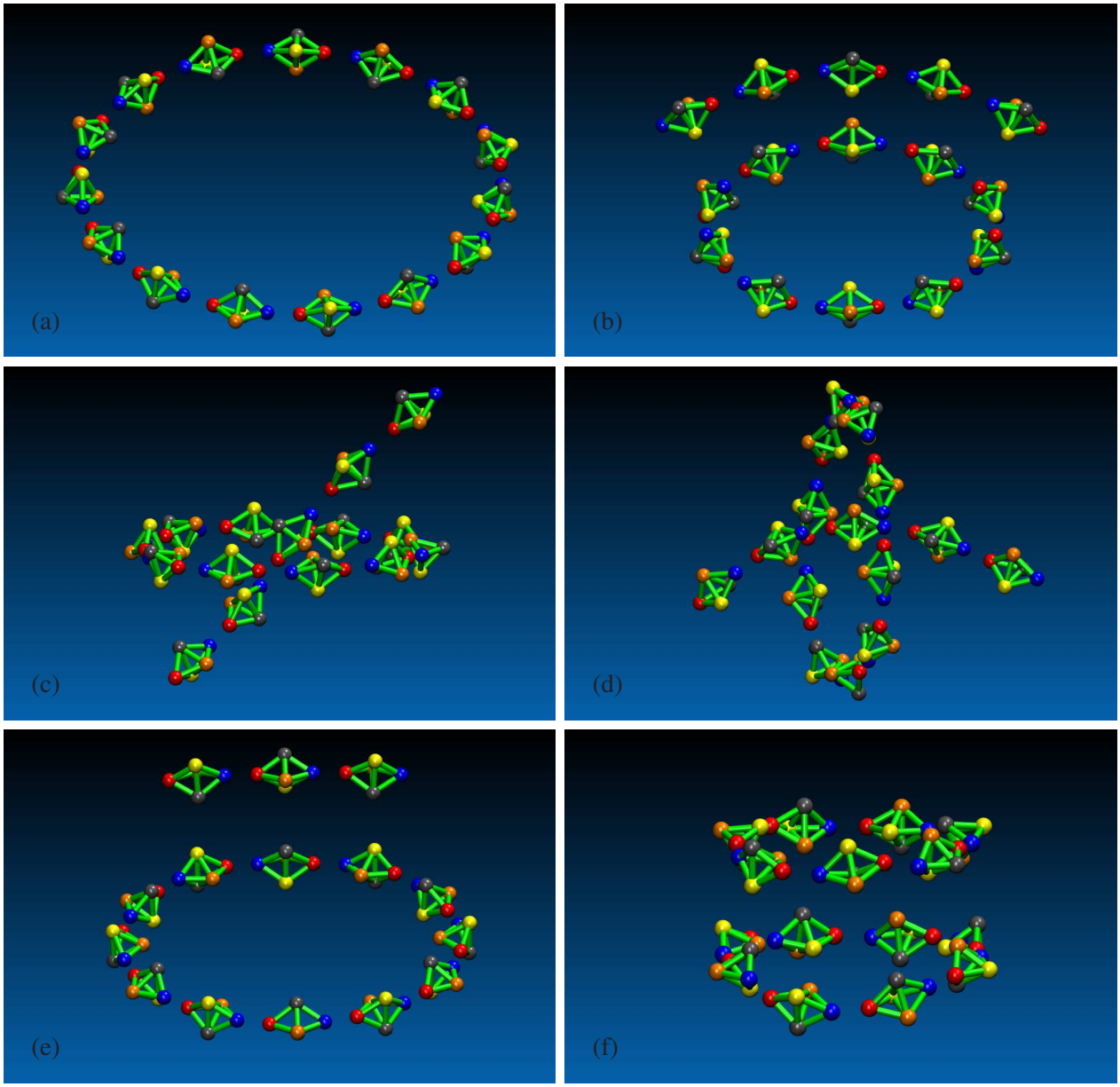

FIG. 9. Some of the structural motifs observed for 15 addressable LJ_5^* clusters. Panels (a) to (e) are the five lowest-lying minima for $\epsilon^{\text{rep}} = 8000$. (a) 15-membered ring, (b) 10-membered ring plus 5-membered chain, (c) and (d) are slightly different rotaxane-like structures, where a five-membered chain passes through a ten-membered ring. (e) 12-membered ring plus 3-membered chain. (f) Stacked 8- and 7-membered rings for $\epsilon^{\text{rep}} = 8000$. All these minima correspond to unconstrained relaxation.

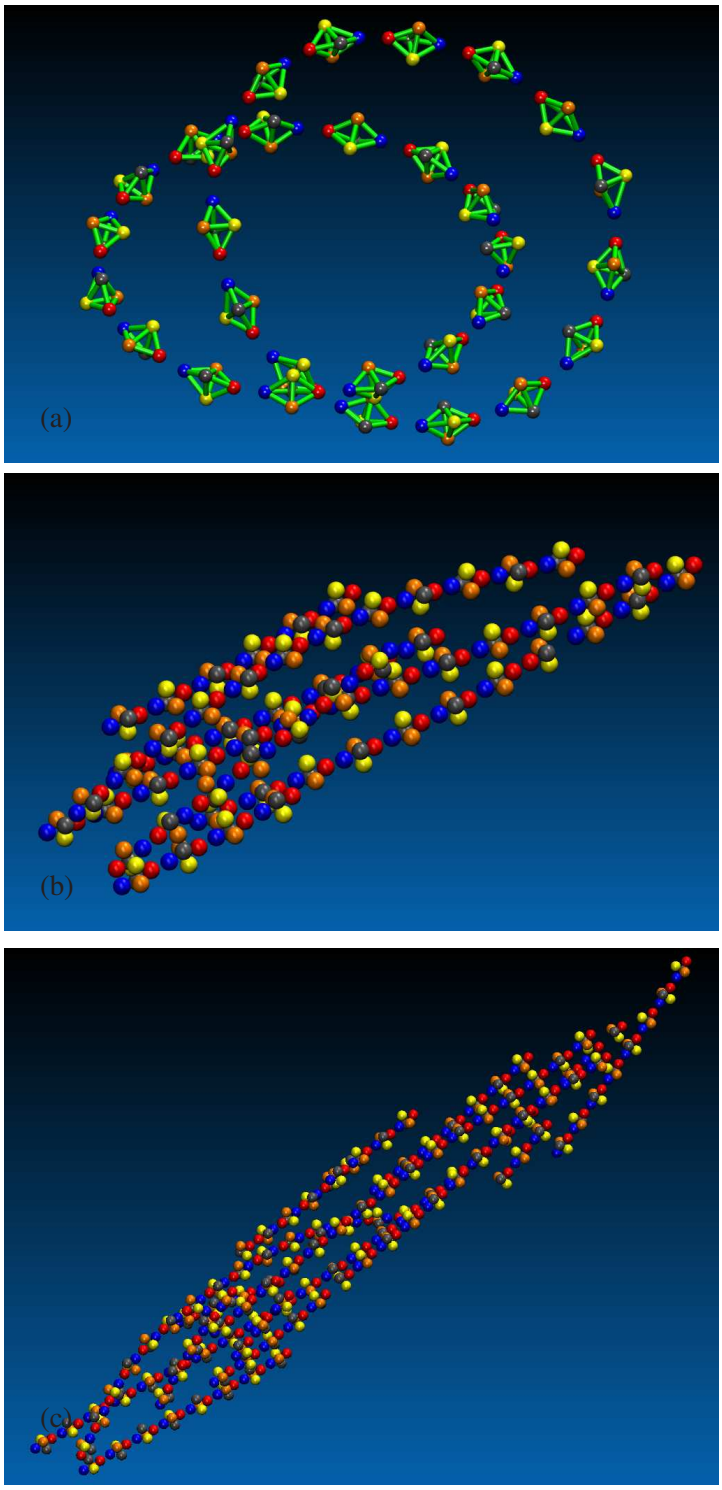

FIG. 10. Selected minima for aggregates of LJ_5* clusters with $\epsilon^{\rm rep} = 15000$. (a) two interpenetrating 15-membered rings, (b) lowest minimum located for 60 clusters, (c) lowest minimum located for 120 clusters. The minima illustrated in panels (b) and (c) result from short basin-hopping runs, and should be representative of low-lying structures. The catenane structure in (a) was seeded from two 15-cluster rings; it is one of the lowest energy minima located in short surveys of the landscape for 30 LJ_5* clusters. All these minima correspond to unconstrained relaxation. To produce a clearer visualisation for the larger aggregates in panels (b) and (c) the atom size is increased and nearest-neighbour edges are omitted.

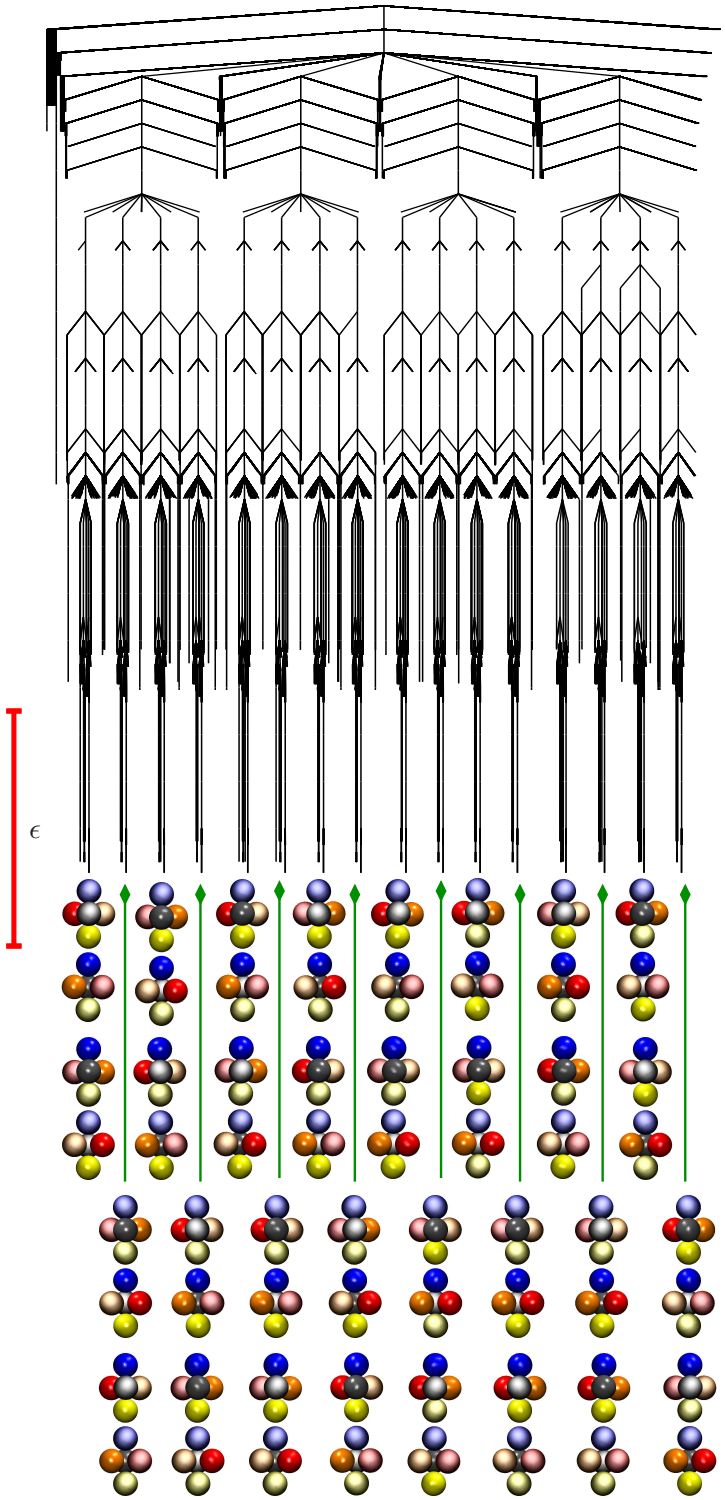

FIG. 11. Disconnectivity graph for the LJ_5* dimer with $\epsilon^{rep} = 8000$. The 32 distinct permutational isomers appear in pairs at the bottom of the 16 subfunnels. The equivalent pairs of atoms are coloured bright and pale shades of blue, red, orange, silver and yellow, using a consistent numbering scheme.

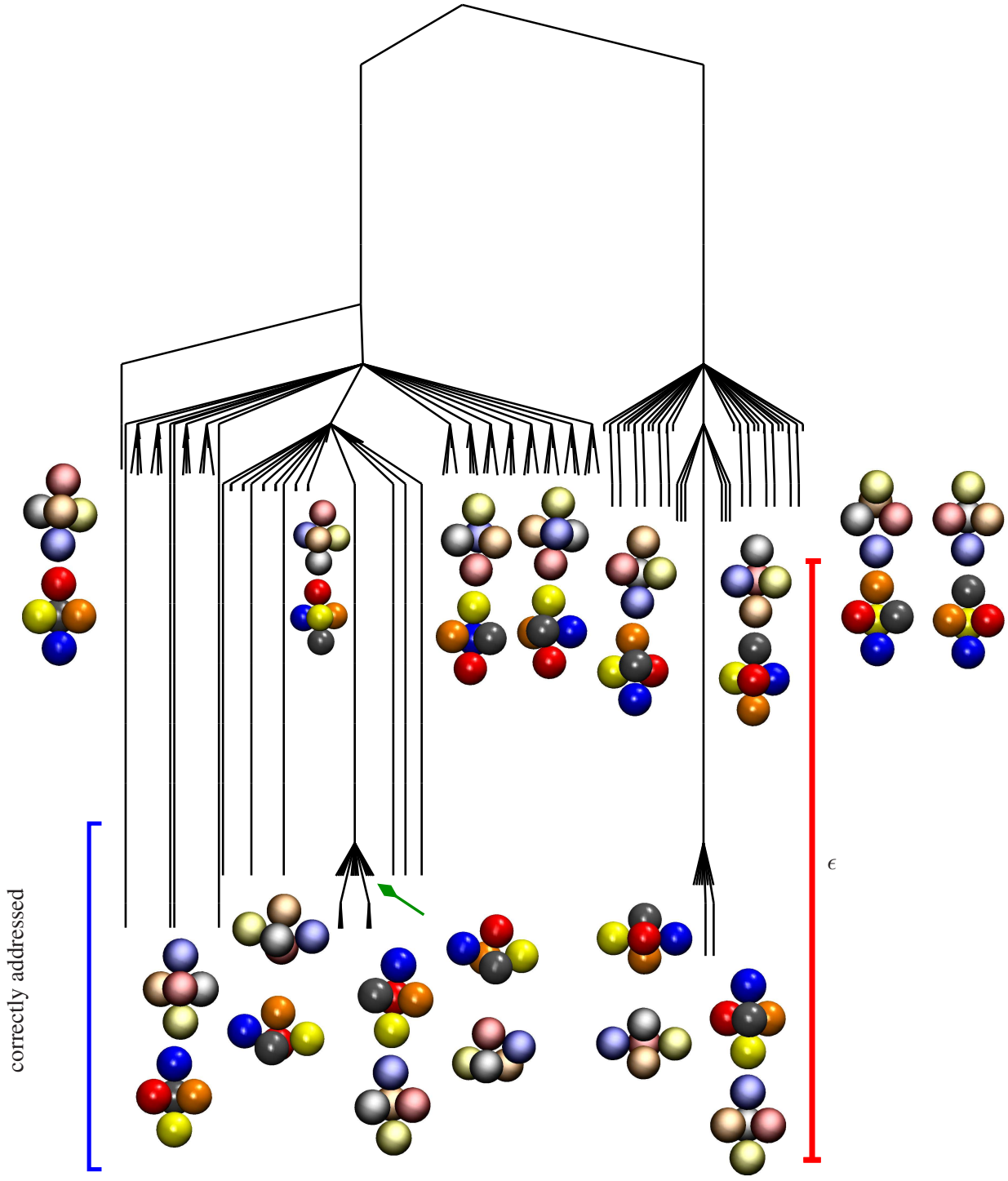

FIG. 12. Disconnectivity graph for one subfunnel of the LJ_5* dimer with $\epsilon^{rep} = 8000$. Selected minima are illustrated close to the corresponding branches. The equivalent pairs of atoms are coloured bright and pale shades of blue, red, orange, silver and yellow, using a consistent numbering scheme. Dimers composed of distinguishable correctly addressed target clusters are well separated from minima with address errors or distortions of the target geometry.